面向新闻推荐的用户兴趣模型构建与更新

袁仁进，陈　刚，李　锋

(信息工程大学 地理空间信息学院，郑州 450001)

摘　要：针对新闻推荐系统中用户兴趣模型构建与用户兴趣漂移问题，提出了一种面向新闻推荐的用户兴趣模型构建与更新方法。首先，采用了向量空间模型与 Bisecting K-means 聚类算法构建了原始用户兴趣模型；然后，以艾宾浩斯遗忘曲线为基础构造了遗忘函数，并以此对用户兴趣模型进行时间加权，从而达到对用户兴趣模型更新的目的。实验以基于用户的协同过滤推荐、基于物品的协同过滤推荐为 baseline，实验结果表明所构建的原始用户兴趣模型推荐性能更优，在 F 值上提升了 4%，更新后的模型比原始模型相比 F 值提高了 1.3%。

关键词：个性化推荐；向量空间模型；用户兴趣模型；用户兴趣漂移；遗忘函数

中图分类号：TP391.3　　doi: 10.3969/j.issn.1001-3695.2018.05.0388

User interest model construction and update for news recommendation

Yuan Renjin, Chen Gang, Li Feng

(*Institute of Geospatial Information*, *Information Engineering University*, *Zhengzhou* 450001, *China*)

Abstract: Aiming at the problem of user interest model construction and user interest drift in news recommendation system, this paper proposed a method of constructing and updating user interest model for news recommendation. Firstly, this method utilized vector space model and Bisecting K-means clustering algorithm to construct the user interest model. Then, it constructed the forgetting function based on Ebbinghaus's forgetting curve, and utilized the forgetting function to update the user interest model by adding time-weight. In the experiments, this paper took user-based and item-based collaborative filtering recommendation as baseline, and the result shows that the original model proposed in this paper has better recommendation performance, with an increase of 4% in F-value, and the updated model is higher than the original model in F-value by 1.3%.

Key words: personalized recommendation; vector space model; user interest model; user interest drift; forgetting function

在信息化、大数据时代，面对数据量的爆炸式增长，个性化推荐技术已经成为各个领域有效利用海量资源信息为用户提供个性化服务的一种首选方案。目前已经在电子商务、音乐、新闻、电影等领域发挥着重要作用。

用户兴趣建模是推荐系统的关键技术之一。协同过滤推荐算法[1,2]作为推荐系统中的经典算法，已被诸多学者应用在新闻推荐中，并在协同过滤推荐算法基础上构建了用户兴趣模型[3~5]。但基于协同过滤的算法并未考虑新闻内容并存在可解释性差和数据稀疏问题，Okura 等人错误!未找到引用源。考虑了新闻内容和用户偏好，使用递归神经网络（RNN）构建了用户兴趣模型，取得了较好效果；Zhang 等人错误!未找到引用源。为缓解文本不足，提出了结合矩阵分解、主题分析和指示图表示的协同模型。同时新闻内容和新闻分类会对推荐效果产生较大影响，在新闻分类方面，古万荣等人错误!未找到引用源。和李佳珊错误!未找到引用源。对新闻聚类算法进行了总结和分析，但未对用户兴趣漂移进行研究。

在实际情况中，新闻具有很强的时效性，同时用户的兴趣也会随着时间变化而漂移。目前针对用户兴趣漂移算法主要有时间窗口法、遗忘函数法以及混合算法三种。时间窗口法是利用时间窗口的移动筛选出用户的最新兴趣，文献**错误!未找到引用源。**对该方面进行了研究。遗忘函数法是利用遗忘函数改变用户不同时间感兴趣物品的权重，成伟丹错误!未找到引用源。在协同过滤算法基础上，采用艾宾浩斯遗忘曲线来描述用户兴趣漂移；Sun 等人错误!未找到引用源。基于聚类和最近邻构建了用户兴趣漂移的动态模型。混合算法是前不同算法的有机混合。邢春晓等人错误!未找到引用源。以协同过滤算法为基础，提出了一种混合算法对用户兴趣变化问题进行了探讨。但这些方法有的仅研究用户兴趣漂移问题，有的是在协同过滤算法基础上研究用户兴趣漂移问题，对解决新闻推荐领域问题还有待研究。

收稿日期：2018-05-03；**修回日期：**2018-06-19

作者简介：袁仁进（1994-），男，湖北鄂州人，硕士研究生，主要研究方向为个性化推荐、数据挖掘、战场环境数据分析（1104372469@qq.com）；陈刚（1971-），男，湖南凤凰人，教授，博导，博士，主要研究方向为虚拟地理环境、战场环境建模与仿真、个性化推荐；李锋（1982-），男，四川邻水人，副教授，讲师，博士，主要研究方向为战场环境仿真、个性化推荐.

针对上述问题，本文提出了一种面向新闻推荐的用户兴趣模型构建与更新方法。首先基于向量空间模型（vector space model，VSM）和 Bisecting K-means 聚类构建用户兴趣模型。由于艾宾浩斯遗忘曲线更符合人的记忆规律，所以接着利用艾宾浩斯遗忘曲线构造遗忘函数来研究用户兴趣模型更新问题。该方法既保留了基于内容的推荐算法的可解释性的优势，同时使用聚类算法避免了编辑分类存在的问题，并对用户兴趣模型更新进行了研究，从而能更好地反映用户兴趣变化情况，提升新闻推荐的准确度。

1　相关工作和流程框架

传统构建新闻推荐系统大多基于协同过滤推荐算法，根据用户浏览新闻纪录或者对新闻的反馈行为来构建用户兴趣模型，达到推荐的目的。此外，有些学者也考虑了新闻分类的情况，但是新闻类别是根据编辑分类得出的，分类标准只能代表编辑的意见，从而导致推荐效果会有所偏差。本文中，为考虑新闻内容和新闻类别对用户兴趣的重要影响，应当确定以下两点工作：a)选取表征新闻内容的模型；b)对已向量化的新闻分类。

a)选取表征新闻内容的模型。目前主流的文本特征表示模型主要包括布尔模型、概率检索模型、语言模型和向量空间模型[错误!未找到引用源。]四种。考虑到新闻数据的高维性，以及为便于新闻聚类从而构建用户兴趣模型，因此将采用向量空间模型来表示新闻特征向量。

b)对已向量化的新闻分类。目前在数据挖掘领域的聚类算法主要包括基于模型的算法、基于网格的算法、基于密度的算法、基于距离的算法四种[错误!未找到引用源。]。本文研究的数据为新闻数据，具有海量、高维等特点，同时采用了向量空间模型来表示新闻的文本特征，因此基于以上考虑，本文采用基于距离的算法作为新闻聚类方法。Abuaiadah [错误!未找到引用源。] 和 Yu [错误!未找到引用源。]研究得出 K-means 聚类算法的改进算法 Bisecting K-means 聚类算法收敛速度更快、聚类效果更优。综上原因，本文将在向量空间模型基础上，采用 Bisecting K-means 聚类算法来实现新闻的分类，具体原理参考文献[15,16]，本文就不详细介绍。

本文提出的用户向量模型构建与更新流程如图 1 所示。首先对以下相关概念进行描述：

a)新闻关键词。新闻中最具有代表性的词，能表征新闻的唯一性和独特性，通常采用文本处理算法进行提取。

b)新闻特征向量。由于新闻所包含的内容属于文本类型，所以使用一个多维向量（$d=(w_1,w_2,\dots,w_m)$）来表示新闻的内容，将新闻文本向量化的结果称为新闻特征向量。

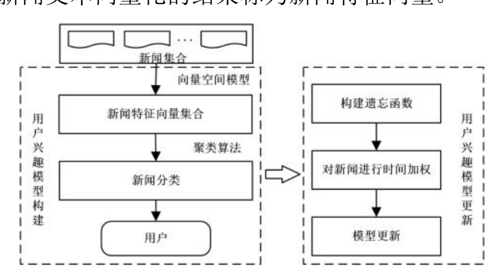


图 1　用户兴趣模型构建与更新框架

2　用户兴趣模型构建

2.1　新闻文本向量化

对于新闻集$D=\{d_1,d_2,\dots,d_n\}$，新闻集的 VSM 表示为

$$M=\begin{bmatrix} w_{11} & \cdots & w_{1m} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nm} \end{bmatrix} \tag{1}$$

$[w_{i1},w_{i2},\dots w_{im}]$表示新闻$i$的新闻特征向量，其中$w_{ij}$表示为关键词$j$在新闻$i$中的权重。构建 VSM 的关键有两个方面：一是确定关键词集的维度m；二是权重w_{ij}的计算。

a)关键词集的维度m。首先提取出每篇新闻的关键词，接着采用 TF-IDF 算法得出新闻集中关键词集的维度m。

b)权重w_{ij}的计算。最常用和有效的权重的计算方法为 TF-IDF 表示法。该方法是信息检索领域的成熟技术，本文就不详细展开。

为使权重值处于[0，1]区间内且新闻能够用等长向量表示，使用余弦归一化的方式对权重进行归一化处理，权重计算公式为

$$w_{(i,j)}=\frac{\mathrm{TF-IDF}(i,j)}{\sqrt{\sum_{i=1}^{T}\mathrm{TF-IDF}(i,j)^2}} \tag{2}$$

2.2　构建基于层次结构的用户兴趣模型

本文中，通过用户已经浏览过的新闻数据构建用户兴趣模型，用户兴趣模型采用三层层次结构表示：用户—新闻类别—新闻。如图 2 所示，第一层节点为用户，第二层节点为用户浏览的新闻类别，第三层节点为用户浏览过的新闻。

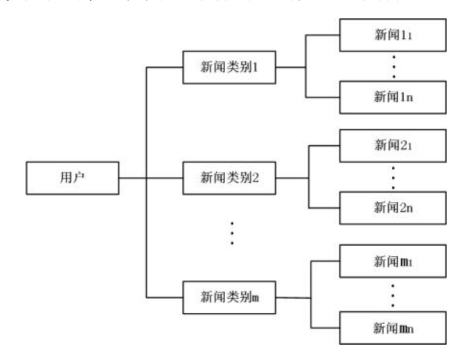


图 2　基于层次结构的用户兴趣模型

若用户浏览过m个不同的新闻类别，则用户兴趣模型可用如下模型表示：

$$user=\{(\boldsymbol{T}_1,w_1,n_1),(\boldsymbol{T}_2,w_2,n_2),\dots,(\boldsymbol{T}_m,w_m,n_m)\} \tag{3}$$

其中：$\boldsymbol{T}_i$表示第i个新闻类别的特征向量；w_i表示第i个新闻类别的权重；n_i表示第i个新闻类别包含的用户浏览过的新闻的数量。

某个新闻类别的特征向量根据该类别所包含的所有已浏览过的新闻特征向量根据兴趣度加权平均求出，即第i个新闻类别的特征向量$\boldsymbol{T}_i$的计算公式为

$$\boldsymbol{T}_i=\frac{\sum_{e_j\in E_j}\boldsymbol{e}_j\cdot I_j}{\sum_{e_j\in E_j}I_j} \tag{4}$$

其中：E_j表示新闻类别i中的用户浏览过的新闻集合；$\boldsymbol{e}_j$表示新

闻特征向量；I_j表示为该类别中第j个新闻的用户兴趣度，用户浏览过某新闻即表示用户对该新闻有兴趣，因此将I_j设为 1，则式（4）可化简为

$$\boldsymbol{T}_i = \frac{\sum_{e_j \in E_j} \boldsymbol{e}_j}{n_i} \tag{5}$$

w_i的值根据第i个新闻类别中用户浏览过的新闻数量占该用户总共浏览过的新闻数量的权重来计算，如式（6）所示。

$$w_i = \frac{n_i}{\sum_{i=1}^{m} n_i} \tag{6}$$

在计算时，用户兴趣模型表示为

$$\boldsymbol{V}_{user} = (w_1 * \boldsymbol{T}_1, w_2 * \boldsymbol{T}_2, \dots, w_m * \boldsymbol{T}_m)^T \tag{7}$$

最终，使用余弦相似度计算候选新闻d_i与用户之间的相似性，计算公式如（8）所示。

$$sim(user, \boldsymbol{V}_{d_i}) = cos(w_i * \boldsymbol{T}_i{}^T, \boldsymbol{V}_{d_i}) \tag{8}$$

其中：$w_i * \boldsymbol{T}_i{}^T$为候选新闻$d_i$所属新闻类别的特征向量；$\boldsymbol{V}_{d_i}$为$d_i$的特征向量。

3　用户兴趣模型更新

3.1　遗忘函数

德国著名心理学家艾宾浩斯错误!未找到引用源。的研究指出：人的遗忘规律不是线性变化的，而是随着时间呈现非线性变化的趋势。在推荐系统中，用户的兴趣也是随着时间变化而改变的，而且应当符合遗忘规律，越早浏览过的新闻在脑海中的印象越低，其在用户兴趣模型中所占的权重也就越小。

根据此规律，构造与艾宾浩斯遗忘曲线类似的基于时间的遗忘函数，其定义如式（9）所示。

$$F(u,i) = a + (1-a)e^{-\frac{T_{ui}}{T_u}} \tag{9}$$

其中：T_{ui}：表示用户u浏览的新闻i与用户最近浏览新闻的时间间隔；T_u：表示用户最早浏览某新闻与最近浏览某新闻的时间间隔；a：表示权重调控因子，$a \in [0,1]$。

图 3 为艾宾浩斯遗忘曲线，展示了记忆保存比率随时间变化的关系。图 4 为本文构造的遗忘函数变化趋势图，展示了用户对新闻的记忆保存比率与新闻浏览时长间隔之间的关系，同时a值的变化会影响遗忘函数的变化趋势。根据不同的推荐系统，可动态调整a值达到最优推荐性能。

3.2　基于遗忘函数的用户兴趣模型更新

用户浏览的新闻在用户脑海中的印象会随着时间推移而减弱，本文采取式（9）中的遗忘函数来量化用户已浏览的新闻在用户中的权重，即表示新闻在用户中的权重会随着时间发生变化。在用户u中，新闻i的权重t_{ui}为

$$t_{ui} = F(u,i) \tag{10}$$

将遗忘函数引入到上文提出的用户兴趣模型中，更新用户兴趣模型。对于用户已浏览的新闻j，其所含内容由<$\boldsymbol{e}_j$>变化为<$\boldsymbol{e}_j, t_{uj}$>，在式（5）中加入新闻权重得到更新后的第i个新闻类别的特征向量$\boldsymbol{T}_i{}'$。

$$\boldsymbol{T}_i{}' = \frac{\sum_{e_j \in E_j} \boldsymbol{e}_j * t_{uj}}{n_i} \tag{11}$$

更新的用户兴趣模型由式（7）变为式（12）。

$$\boldsymbol{V}_{user}{}' = (w_1 * \boldsymbol{T}_1{}', w_2 * \boldsymbol{T}_2{}', \dots, w_m * \boldsymbol{T}_m{}')^T \tag{12}$$

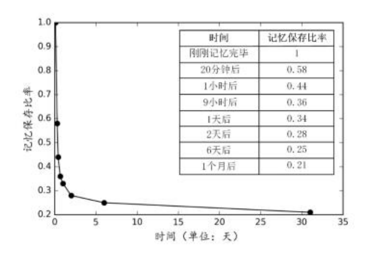


图 3　艾宾浩斯遗忘曲线

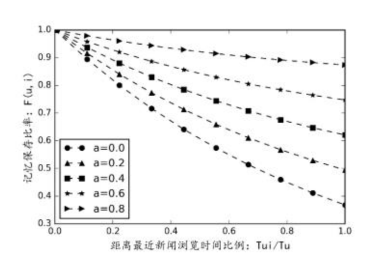


图 4　a值对遗忘函数变化趋势影响

4　实验及对比分析

4.1　数据准备

本文使用 DataCastle 提供的用户浏览新闻记录作为实验数据集，该数据集从国内财新网随机采集，总共包括 10 000 名用户在 2014 年 3 月期间的浏览的 116 225 条记录。每条浏览记录包括用户编号、新闻编号、浏览时间以及新闻文本内容等。实验中，采用 python 的第三方库——jieba 分词器进行分词，根据实际新闻内容采用改进的哈尔滨工业大学信息检索中心的停用词表去除停用词。

首先对原始数据进行预处理，将浏览记录超过 30 条的用户筛选出，共得到 417 名用户的 32 770 条浏览记录。将这些数据随机分成 5 组，每组包含 200 名用户，并将每个用户的后 10 条浏览记录作为测试集。每组数据中的抽样结果有重复，为保证实验结果客观性，将 5 组数据的测试结果取平均值作为最后的实验结果。

4.2　评估指标

为平衡准确率与召回率两者之间的结果，本文采用 F 值作为实验的评估指标，F 值由准确率和召回率组合计算。准确率和召回率由混淆矩阵表示，如表 1 所示。

表 1　混淆矩阵

	被推荐	未被推荐
喜欢	true positive(TP)	false negative(FN)
不喜欢	false positive(FP)	true negative(TN)

准确率P、召回率R的计算公式为

$$P = \frac{TP}{TP+FP} \tag{13}$$

$$R = \frac{TP}{TP+FN} \tag{14}$$

F 值的计算使用准确率P、召回率R共同表示为

$$F = \frac{2PR}{P+R} \tag{15}$$

4.3 实验结果与分析

为验证本文所提方法的可行性以及推荐效果，实验中以基于用户的协同过滤推荐算法错误!未找到引用源。、基于物品的协同过滤推荐算法错误!未找到引用源。为 Baseline 进行对比分析，对这两种协同过滤推荐算法考虑的最近邻数包括（5,10,15,20）四种，并考虑五种不同的推荐结果个数（10,15,20,25,30）。

为研究新闻集的聚类簇数M对推荐性能产生的影响，Bisecting K-means 聚类算法考虑的聚类簇数M包括（10,15,20,25,30）五种，对应的评估指标结果如表 2 所示。

表 2　聚类簇数 M 对本文算法 F 值影响

簇数M	$N=10$	$N=15$	$N=20$	$N=25$	$N=30$
10	F=0.164	F=0.175	F=0.180	F=0.173	F=0.162
15	F=0.182	F=0.196	F=0.205	F=0.192	F=0.187
20	F=0.193	F=0.211	F=0.215	F=0.187	F=0.179
25	F=0.175	F=0.197	F=0.201	F=0.198	F=0.183
30	F=0.171	F=0.179	F=0.185	F=0.178	F=0.169

观察表 2 结果可知，当推荐结果个数不变时，Bisecting K-means 聚类算法中的聚类簇数M与评估指标 F 值有非线性关系，具体表现为 F 值随着M的增加呈现先增大后减小的变化趋势，并当$M=20$时，F 值最大，即推荐性能最优。

将每种方法中不同情况的实验结果进行比较，选取最优的结果作为每种方法的实验数据。主要包括两个实验：a）本文所构建的模型与协同过滤算法相比较；b）更新后的模型与之前模型相比较。

1）本文所构建的模型与协同过滤算法相比较

将实验得出的五组 F 值数据计算平均值，不同算法得出的 F 值平均值如表 3 所示。

表 3　本文所构建的模型与协同过滤算法的 F 值相比较

算法	$N=10$	$N=15$	$N=20$	$N=25$	$N=30$
基于用户	F=0.166	F=0.173	F=0.168	F=0.157	F=0.148
基于物品	F=0.162	F=0.174	F=0.164	F=0.160	F=0.146
本文模型	F=0.193	F=0.211	F=0.215	F=0.187	F=0.179

图 5 展示了表 3 中不同算法的 F 值变化情况。从图中可知本文模型的推荐效果与两种协同过滤推荐算法相比有较大提升，推荐效果更优。当推荐新闻个数在[15，20]区间内时，各种算法的推荐效果相对来说最好，在该区间内，本文模型的 F 值比两种协同过滤算法平均高于 4%。

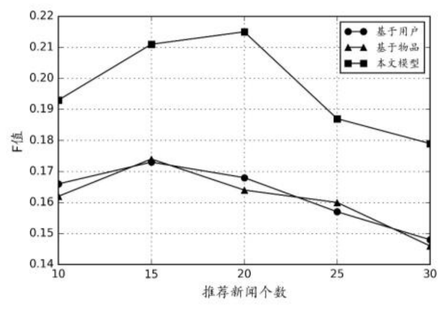


图 5　不同算法之间 F 值比较

2）更新后的模型与之前模型相比较

更新后的模型推荐性能受式（9）中遗忘函数的权重调控因子a所影响。为了解a值与推荐性能之间的关系，实验中考虑的a值为（0,0.2,0.4,0.6,0.8）共五种情况。表 4 展示了更新后模型中不同的a值在不同推荐结果个数中的推荐性能。

表 4　a 值对推荐性能影响

更新后模型	$N=10$	$N=15$	$N=20$	$N=25$	$N=30$
$a=0$	F=0.185	F=0.191	F=0.193	F=0.183	F=0.172
$a=0.2$	F=0.192	F=0.209	F=0.211	F=0.183	F=0.174
$a=0.4$	F=0.203	F=0.221	F=0.223	F=0.210	F=0.194
$a=0.6$	F=0.202	F=0.219	F=0.220	F=0.203	F=0.189
$a=0.8$	F=0.173	F=0.185	F=0.186	F=0.169	F=0.157

图 6 展示了表 4 中a值变化引起推荐性能改变的变化趋势。可知，当$a=0.4$时，更新后模型的推荐性能最优。在$a=0.4$的遗忘函数下，更新后的模型与之前的模型进行比较，两者之间的推荐性能对比如图 7 所示。更新后模型的性能整体优于之前模型，F 值平均提高 1.3%，说明用户兴趣确实随着时间变化而漂移；同时本文提出的遗忘函数有助于用户兴趣模型更新，但效果不是很明显，探讨遗忘函数对推荐性能的影响还需进一步研究。

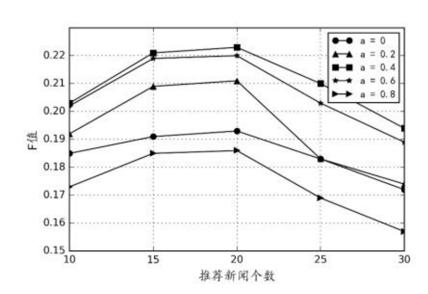


图 6　a 值对更新后模型的推荐性能影响

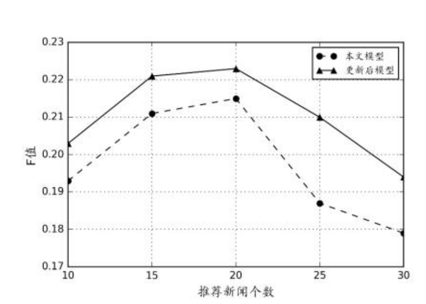


图 7　本文模型与更新后的模型比较

5　结束语

a)与基于用户的协同过滤推荐算法和基于物品的协同过滤推荐算法相比，本文结合 VSM 和 Bisecting K-means 聚类算法构建的用户兴趣模型推荐效果更好，在 F 值指标上平均提高了 4%。

b）利用遗忘函数更新后的用户兴趣模型与原始模型相比，推荐性能稍有提升，在 F 值指标上平均提高了 1.3%。

c）在实验数据中，F 值一直呈现先高后低的趋势，这一现象本文目前还难以解释，还需进一步深入研究。

本文针对新闻内容和新闻分类会对新闻推荐系统的推荐性能产生影响的情况下，首先提出了一种基于 VSM 和 Bisecting K-means 聚类的用户兴趣模型构建方法；考虑到用户兴趣会随着时间而发生漂移，因此在该模型基础上，参考艾宾浩斯遗忘曲线构造了遗忘函数，并利用遗忘函数对该模型进行了时间加权更新。实验结果表明，本文构建的模型与协同过滤算法相比，推荐性能有所提高；更新后的模型与原始模型相比，推荐性又有进一步改善，但提升较少。总的来看，本文提出的方法可作为新闻推荐领域的一种方案，在模型构建与利用遗忘函数更新模型方面可作为研究参考。在本文基础上，下一步将进一步研究用户兴趣模型更新问题，以及对结合地理位置信息的新闻推荐方法进行研究。

参考文献：

[1] 冷亚军，陆青，梁昌勇. 协同过滤推荐技术综述 [J]. 模式识别与人工智能, 2014, 27 (8): 720-734. (Leng Yajun, Lu Qing, Liang Changyong. Survey of recommendation based on collaborative filtering [J]. PR&AI, 2014, 27 (8): 720-734.)

[2] 孔艳莉. 基于协同过滤算法的个性化推荐技术研究 [D]. 北京：北京工业大学, 2016. (Kong Yanli. Research on personalized recommendation based on collaborative filtering [D]. Beijing: Beijing University of Technology, 2016.)

[3] 吴彦文，齐旻，杨锐. 一种基于改进型协同过滤算法的新闻推荐系统 [J]. 计算机工程与科学, 2017, 39 (6): 1179-1185. (Wu Yanwen, Qi Min, Yang Rui. A new recommendation system based on an improved collaborative filtering algorithm [J]. Computer Engineering & Science, 2017, 39 (6): 1179-1185.)

[4] Garcin F, Zhou Kai, Faltings B, *et al.* Personalized news recommendation based on collaborative filtering [C]// Proc of IEEE//WIC//ACM International Conferences on Web Intelligence and Intelligent Agent Technology. [S. l.] : IEEE Press, 2013: 437-441.

[5] 彭菲菲，钱旭. 基于用户关注度的个性化新闻推荐系统 [J]. 计算机应用研究, 2012, 29 (3): 1005-1007. (Peng Feifei, Qian Xu. Personalized recommendation system for news based on user concern [J]. Application Research of Computers, 2012, 29 (3): 1005-1007.)

[6] Okura S, Tagami Y, Ono S, *et al.* Embedding-based news recommendation for millions of users [C]// Proc of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S. l.] : ACM Press, 2017: 1933-1942.

[7] Zhang Kuai, Xin Xin, Luo Pei, *et al.* Fine-grained news recommendation by fusing matrix factorization, topic analysis and knowledge graph representation [C]// Proc of IEEE International Conference on Systems, Man and Cybernetics. [S. l.] : IEEE Press, 2017: 918-923.

[8] 古万荣，董守斌，何锦潮，等. 基于二次聚类的新闻推荐方法 [J]. 华南理工大学学报：自然科学版, 2014, 42 (7): 15-20, 32. (Gu Wanrong, Dong Shoubin, He Jingchao, *et al.* News recommendation method based on secondary clustering [J]. Journal of South China University of Technology: Natural Science, 2014, 42 (7): 15-20+32.)

[9] 李佳珊. 个性化新闻推荐引擎中新闻分组聚类技术的研究与实现 [D]. 北京：北京邮电大学, 2013: 20-29. (Li Jiashan. Research and implementation of text clustering for personalized news recommendation system [D]. Beijing: Beijing University of Posts and Telecommunications, 2013: 20-29.)

[10] 费洪晓，戴弋，穆珺，等. 基于优化时间窗的用户兴趣漂移方法 [J]. 计算机工程, 2008, 34 (16): 210-211. (Fei Hongxiao, Dai Yi, Mu Jun, *et al.* Method of drifting user’s interests based on time window optimization [J]. Computer Engineering, 2008, 34 (16): 210-211.)

[11] 成伟丹. 基于遗忘函数和用户的协同过滤推荐算法研究 [D]. 杭州：浙江工业大学, 2016. (Cheng Weidan. Research on collaborative filtering algorithm based on forgetting function and user [D]. Hangzhou: Zhejiang University of Technology, 2016.)

[12] Sun Baoshan, Dong Lingyu. Dynamic model adaptive to user interest drift based on cluster and nearest neighbors [J]. IEEE Access, 2017, PP (99): 1.

[13] 邢春晓，高凤荣，战思南，等. 适应用户兴趣变化的协同过滤推荐算法 [J]. 计算机研究与发展, 2007, 44 (2): 296-301. (Xing Chunxiao, Gao Fengrong, Zhan Sinan, *et al.* A collaborative filtering recommendation algorithm incorporated with user interest change [J]. Journal of Computer Research and Development, 2007, 44 (2): 296-301.)

[14] 姚清耘. 基于向量空间模型的中文文本聚类方法的研究 [D]. 上海：上海交通大学, 2008: 27. (Yao Qingyun. Research of VSM-based Chinese text clustering algorithms [D]. Shanghai: Shanghai Jiao Tong University, 2008: 27.)

[15] Abuaiadah D. Using bisect K-means clustering technique in the analysis of Arabic documents [J]. ACM Trans on Asian and Low-Resource Language Information Processing, 2016, 15 (3): 1-13.

[16] Yu Zhuang. Symmetric repositioning of bisecting K-means centers for increased reduction of distance calculations for big data clustering [C]// Proc of IEEE International Conference on Big Data. [S. l.] : IEEE Press, 2017: 2709-2715.

[17] Ebbinghaus H. Memory: a contribution to experimental psychology [J]. Annals of Neurosciences, 2013, 20 (4): 155.

[18] 项亮. 推荐系统实践 [M]. 北京：人民邮电出版社, 2012: 44-59. (Xiang Liang. Recommended system practice [M]. Beijing: Post & Telecom Press, 2012: 44-59.)